Returns to Initial Years of Formal Education: How Birthdate Affects Later Educational Outcomes

Asaad Ismail Ali
asaad.ali@pg.canterbury.ac.nz
&
Andrea Kutinova Menclova
andrea.menclova@canterbury.ac.nz

Department of Economics and Finance
University of Canterbury

Abstract

In most developed countries, children start school on a fixed date (e.g., early September) in contrast to New Zealand where there are rolling admissions and children can start school right after their 5th birthday.

If a child's birth date is between January and May, the young New Zealander will typically spend the year he/she turns 5 in Year 1 and the next year in Year 2. If a child's birth date is between June and December, the student will usually spend the year he/she turns 5 in Year 0 and the next year in Year 1. This means that the date of birth affects the amount of time spent in primary school and may further result in different educational outcomes.

In this paper, we analyse the effects of school start on long-term educational achievement in New Zealand. Specifically, we focus on National Certificate of Educational Achievement (NCEA) and University Entrance (UE) results. Controlling for demographic and socio-economic characteristics, we find that an additional month spent in Years 0/1 increases the probability of achieving NCEA level 1 by 2%, NCEA level 2 by 4%, NCEA level 3 by 6%, and UE by 5%. Thus, differences in the timing of birth – and hence in school start – seem to have large effects on achievement even years later, in high school.

Keywords: Returns to education, school start, achievement

JEL codes: I21, I26

Statistics New Zealand Disclaimer

The results in this paper are not official statistics. They have been created for research purposes from the Integrated Data Infrastructure (IDI), managed by Statistics New Zealand (Stats NZ).

The opinions, findings, recommendations, and conclusions expressed in this thesis are those of the authors, not Statistics NZ, IDI or Ministry of Education.

Access to the anonymised data used in this study was provided by Statistics NZ under the security and confidentiality provisions of the Statistics Act 1975. Only people authorised by the Statistics Act 1975 are allowed to see data about a particular person, household, business, or organisation, and the results in this paper have been confidentialised to protect these groups from identification and to keep their data safe.

Careful consideration has been given to the privacy, security, and confidentiality issues associated with using administrative and survey data in the IDI. Further detail can be found in the Privacy impact assessment for the Integrated Data Infrastructure available from www.stats.govt.nz

I. Background

Unlike many other developed countries such as the United States of America, United Kingdom, most of the European Union, and Australia, where primary school attendance starts for all children on a specific date (e.g., in early September), New Zealand schooling officially starts when a child reaches the age of five. Schooling from ages six to 16 is compulsory. School term 1 begins in February and the primary education system goes from Year 0 to 8. If a child's birth date is between January and May, the young student will typically spend the year he/she turns five in Year 1 and the next year in Year 2. If a child's birth date is between June and December, the student will usually spend the year he/she turns five in Year 0 and start Year 1 the following February. This means that the date of birth affects the amount of time spent in primary school and may further result in different educational outcomes.

The main objective of this paper is to make use of the unusual school start policy in New Zealand to study the effects of early school attendance on the individual's later educational outcomes at the end of high school, as measured by National Certificate of Educational Achievement (NCEA) and University Entrance (UE) results. The study finds large positive returns to early schooling.

I.A. Previous Evidence

There are three different aspects of school start that have been examined previously in regards to educational achievement. The first is the effect of *absolute age*; i.e., different children start school at a somewhat different age and hence at a different stage of their cognitive and social development (mechanism A). The second is the variation in *relative age* among children starting school; i.e., some are younger/older than their peers (mechanism B). Finally, there is the causal effect of schooling on educational outcomes (mechanism C).

I.A.1. Absolute and Relative Age Effects (Mechanisms A & B)

A number of studies find better academic achievement among children starting school at an older age. Strøm (2004) uses Norwegian data to explore the relationship between the age at school start and children's achievement towards the end of secondary schooling - holding the date at which school starts constant. Strøm's study shows that younger students have a considerable disadvantage compared to older peers within the same class. The oldest students, born in January, generally score higher in reading tests at 15 to 16 years of age. Compared to the youngest students, born in December, their scores are higher by around 20% of the

standard deviation. Strøm adds that he is unable to propose an alternative school start policy which may eliminate this disadvantage.

Datar (2004) examines the effect of postponing kindergarten admission in the USA on children's academic success. Using instrumental variables based on an exogenous discrepancy in birth dates and kindergarten admission age policies, Datar finds that starting kindergarten a year older considerably improves test scores at kindergarten admission. More importantly, the trajectory of test scores is steeper during the first two years of primary school for older children. Datar also suggests that the advantages of delaying kindergarten admission tend to be considerably higher for at-risk, such as poor and disabled, children.

Kawaguchi (2011) uses a Japanese labour force survey to demonstrate that older students in a school group have superior educational achievement and labour market outcomes compared to their younger peers.

Crawford, Dearden and Greaves (2013) show that the oldest children in a particular academic year in England perform considerably better than the youngest children in national achievement tests until the age of 19. Importantly, this difference is experienced when the students turn 16 and make decisions about continuing further secondary school studies as well as when they turn 19 and make decisions about higher education.

Using USA data, Lubotsky and Kaestner (2016) examine whether children with a high level of cognitive and non-cognitive skills at the start of kindergarten experience higher gains in these skills in subsequent years. They show that older kids in kindergarten score higher than the younger ones on both cognitive and non-cognitive measures of achievement. Their cognitive assessment scores grow quicker during kindergarten and first grade. However, the younger entrants start doing better after the first grade and their scores catch up.

The positive effect of older age at school start is not observed universally. For example, Angrist and Krueger (1992) examine the effects of the age at school start on later academic performance in the USA. To get exogenous variation in the age at school start (and hence causal effects), they use mandatory school attendance laws as an instrumental variable. Unlike previous studies (e.g., DiPasquale, Moule,, & Flewelling, 1980; Warren, Levin, & Tyler, 1986) which used children's primary school test scores as the outcome variable, Angrist et al. argue that a superior measure of academic achievement than aptitude test performance at an early age may be the years of education that a child eventually attains. Their results show that children who enter school older attain relatively less education.

Zhang, Zhong and Zhang (2017) use the China Education Panel Survey to test the effect of school starting age on junior high school academic achievement. Their study shows that a one-year delay in starting school decreases student's cognitive scores in 7^{th} grade by 0.3 standard deviations. They further investigate the mechanisms underlying the relationship between age at entrance and educational outcomes and find that the decrease in scores depends on the accumulation of human capital prior to the start of primary school. In the absence of preschools in China, wealthier parents invest a lot more in their children's pre-school development as compared to poor parents.

The relative age effect is also very common in competitive sports. Musch and Grondin (2001) review a wide variety of sports studies on the relative age effects (RAEs) and confirm that RAEs are a common phenomenon in competitive sports. They suggest that bringing RAEs to the attention of all coaches and team managers in the minor sports system is a necessary first step towards safeguarding equal treatment and unbiased competition among players. Barnsley and Thompson (1988) show RAEs in minor hockey. As younger children are at an earlier stage of development than their larger/stronger team members, they are more likely to experience failure and frustration and hence grow an inferior expectation of themselves as hockey players. Boucher and Mutimer (1994) replicate a series of studies (Barnsley & Thompson, 1988; Barnsley, Thompson, & Barnsley, 1985; Daniel & Janssen, 1987; Grondin, Deshaies, & Nault, 1984) of professional ice-hockey players and, like the original studies, find a strong connection between relative age of the players and their participation and contribution in the sport. Cobley et al. (2009) confirm the presence of RAEs through a meta-analytical review of 38 studies, spanning 1984 to 2007, consisting of 253 independent samples across 14 sports and 16 different countries. Fumarco et al. (2017), on the other hand, find an inverse relative age effect in the North American National Hockey League (NHL); i.e. players born in the last quarter of a calendar year score more and have higher earnings than those born in the first quarter.

It is clear from the above literature that age may play a significant role in sports as well as the academic achievement of students. Hence, it is vital to control for the students' age in our analyses. However, given the fixed school start date in most countries, the above articles cannot i) examine whether gradual admittance into early primary education – at a *given age* – eliminates the effect of a student's date of birth on later educational attainment or ii) study the causal effect of the time spent in school on later outcomes. We turn to these issues below.

I.A.2. Causal Effects of the Length of Schooling (Mechanism C)

There are a few studies that try to estimate the causal effect of time spent in school on educational outcomes, using different identification techniques. Some use data on students of the same age but in different grades, i.e. comparable cognitive skills but a different level of education, while others (like us) use a unique school system that allows students to enter school at a certain age instead of a certain date.

Cahan and Cohen (1989) estimate the effects of both age and time spent in school for over 12,000 students in grades 4 to 6 in Israel. The effect of age is measured as the difference in mean predicted scores between the youngest and the oldest students in a particular grade whereas the effect of time spent in school is measured as the difference in the mean predicted scores between the oldest student in that grade and the youngest student in the higher adjacent grade. The authors conclude that one additional year of schooling increases test scores by 0.30 of a standard deviation. On the other hand, being a year older increases test scores by 0.15 of a standard deviation. Therefore, the effect of an additional year of schooling is on average about twice as large as the effect of being a year older.

Cliffordson and Gustafsson (2008) estimate the effects of both age and schooling on various aspects of intellectual performance in Sweden. They base their analysis on the test scores from military enlistment measuring ‘general visualization ability’, ‘crystallized intelligence’ and ‘fluid ability’ at age 16. The tests occur on different dates throughout the year which gives differences in both age and length of schooling among individuals at the time of the test. The authors find that both schooling and age generally increase performance, with the effect of schooling being considerably higher than the effect of age.

Most relevant for our study, Leuven et al. (2010) evaluate the effect of expanding possibilities for early enrolment at school on later achievement using a novel quasi-experimental strategy. They exploit two distinct features of the Dutch schooling system. One is their rolling admissions policy; i.e. children do not have to wait to start primary school on a particular date, they can start right after their fourth birthday. Second, children with birthdays during or right after school holidays start at the same time (at the beginning of the next term) and are put in the same class. The authors use the exogenous variation created by these distinct features in children’s enrolment opportunities to identify their effects on subsequent test

scores. They conclude that an additional month of schooling for disadvantaged[1] children increases their arithmetic test scores by five percent of a standard deviation and their language test scores by six percent of a standard deviation. The study finds no effects for non-disadvantaged children.

Ali and Menclova (2018) replicate Leuven's study. This replication in general endorses the findings of Leuven et al. but with some notable differences. Specifically, the authors find positive effects of the time spent in school for both disadvantaged and non-disadvantaged children. On average, an additional month of schooling for disadvantaged children increases their arithmetic and language test scores by three percent of a standard deviation. An additional month of schooling for non-disadvantaged children increases their arithmetic test scores by five percent of a standard deviation and their language test scores by four percent of a standard deviation.

For completeness, other studies suggest that early school attendance may have long-term effects beyond academic achievement. For example, Lleras-Muney (2005) shows a large casual effect of education on mortality in the USA. The author estimates the effect using two different ways: GLS and IV estimation. The results from the GLS estimation show that the probability of dying in the next ten years decreases by about 1.3 percentage points with an additional year of education. The IV estimation shows a much larger effect: an additional year of schooling decreases the probability of dying in the next 10 years by about 3.6 percentage points. The study further elaborates on how life expectancy gains can arise from this effect. It shows that in 1960, at age 35, an additional year of education increased the life expectancy by as much as 1.7 years.

I.B. Identification Strategy

As noted above, in New Zealand, the timing of birth – and hence a child's fifth birthday – affects how much time an individual spends in early primary education. If a child is born between January and May, he/she will typically start school in Year 1 and will move to Year 2 the subsequent year. If a child is born between June and December, he/she will likely start school in Year 0 and transition to Year 1 the following February. This means that at the start of Year 2 of primary school, excluding holidays, children's potential time spent in school

[1] Children are classified as disadvantaged if both parents have at most a degree from a low-level vocational school.

varies from approximately 4 to 11 months[2] (refer to Appendix Figure A1 for a graphical exposition). Another important characteristic of the school system is the school holiday period. There are four different school holiday periods in a calendar year. All children born during these holidays start school at the same time on the first day of the new term. Therefore, the amount of time each child can potentially spend in school (maximum length of schooling) varies because of these characteristics and is not a linear function of his/her age. This is key for our identification strategy which follows previous work for the Netherlands by Leuven et al. (2010).

Figure 1 shows the relationship between a child's date of birth and his/her potential 'maximum length of time spent in school'. *Penroll* only includes teaching days while *penroll0* also includes school holidays and weekends. The horizontal segments on *penroll* reflect being born during school holidays and the segments with negative slope are for children born outside of school holidays. There are a total of four horizontal segments reflecting the four periods of school holidays in a calendar year[3]. On our time axis, the first holidays are from July 9th to July 24th, the second from September 24th to October 9th, the third from December 20th to February 6th (which includes the Christmas and New Year holidays), and the fourth from April 14th to April 25th. Children who turn five on the same downward-sloping segment have a one to one relationship between the time potentially spent in school and their age, i.e., an additional day of age leads to an additional day potentially spent in school. Any differences in the test scores of these children can be attributed to changes in their 'maximum length of schooling' as well as changes in their age (or randomly distributed changes in child/parental/regional characteristics). In comparison, children who turn five on the same horizontal segment (i.e., during a holiday period) all start school at the same time after the school holidays in the upcoming school term and so while they differ in age, they do not differ in the maximum time spent in school. Crucially, this allows us to empirically isolate the returns to time spent in school (mechanism C) from relative age effects (mechanism B), while absolute age effects (mechanism A) do not occur in a system where children start school at the same age.

[2] The maximum is 10.8 and the minimum 4.2, leading to a difference of 6.6 months.

[3] These holidays fall on slightly different days each year. We use dates from year 2005 for ease. The total number of holidays (or their length) has not changed over the years.

I.C. New Zealand School System

Broadly speaking, there are three levels of the New Zealand education system:

1. From birth to school entry, known as early childhood education;
2. From Year 0 to Year 13 (about age 5-18), known as primary and secondary education;
3. Above Year 13 (from about age 18 onwards) – higher/tertiary and vocational education.

Our study focuses on stage 2 above, i.e. the effects of early primary education on secondary school achievement and entry into tertiary education. Specifically, we examine the effects of differences in the initial time spent in primary school (due to differences in the dates of birth) on standardised achievement results near the end of high school.

New Zealand secondary schools operate a national qualification system known as the National Certificates of Educational Achievement (NCEA). This is what we use as one of the measures of standardised achievement, as described in detail below.

Another measure to assess the performance of a student, the second measure we use in this study, is known as University Entrance (UE). It is given to students based on specific NCEA results/achievements.

I.C.1. National Certificate of Educational Achievement

The National Certificates of Educational Achievement are the primary national assessment tool for secondary school students in New Zealand (New Zealand Qualifications Authority, 2013/14). The New Zealand Qualifications Authority (NZQA) administers the NCEA and students do not have to apply to participate; they are automatically included.

NCEA qualifications are recognized by businesses, and used by colleges and universities both in New Zealand and abroad. Every student is assigned a unique identifier known as the National Student Number. The student or an employer/university can then use this unique number to search for the individual's NCEA results in an NZQA database.

NCEA tests the performance of students in various subjects, known as *standards*. For example, in mathematics standards, application of numeric thinking is measured. When students demonstrate a required level of knowledge/skills in a standard, they are awarded NCEA *credits*. Students need to obtain a specific number of credits in order to *achieve* an NCEA certification.

NCEA certification has three consecutive levels, based on the level of the evaluated knowledge/skills. Typically, students work through NCEA levels 1-3 in their secondary school Years 11-13, respectively. Receiving NCEA Merit or NCEA Excellence can officially recognize students' quality of work for a given level.

I.C.2. University Entrance

The minimum entrance requirement into a New Zealand university is University Entrance. Gaining UE is the requirement of all New Zealand universities and some universities then have additional requirements beyond UE (Shui, 2017). The UE qualification is based on specific credits from NCEA levels 2 and 3 and is the minimum requirement for direct admission to a university in New Zealand (New Zealand Qualifications Authority, 2013/14).

To qualify for a UE, a student needs:

- An NCEA level 3 qualification;
- Approved subjects: 14 credits in each of three approved subjects[4] at NCEA level 3;
- A literacy requirement: 10 credits at NCEA level 2 or above, made up of 5 credits in reading and 5 credits in writing.
- A numeracy requirement: 10 credits at NCEA level 1 or above in relevant achievement standards; or all three numeracy standards (26623, 26626 and 26627).

Once a student has met the requirements for University Entrance it will appear on his/her Record of Achievement.

I.C.3. School Deciles

School decile is used in our study to proxy for the students' socio-economic background.

Each school in New Zealand has been assigned a decile rating which shows the socio-economic ranking of the census area sending children to each school. Decile 1 schools are the lowest ranked, implying that a high percentage of students in those schools are from a low socio-economic background; decile 10 schools are the highest ranked. By design, each decile contains almost the same number of schools, i.e. roughly 10%. The decile rank is not in any way an indicator of the quality of education provided by the school.

[4] A list of approved subjects is available at: https://www.nzqa.govt.nz/qualifications-standards/awards/university-entrance/approved-subjects/

Historically, the main objective of creating a decile ranking system was to determine how much disadvantage-related funding each state or state-integrated school should get. Schools in low deciles get the most funding per student. The New Zealand Ministry of Education re-calculates deciles every five years. The decile calculation is based on certain relative socio-economic factors of the community that students of a school come from. These factors include: household crowding; percentage of residents with income in the lowest twenty percent nationally; percentage of parents in low-skill occupational groups; percentage of parents without an educational qualification; and percentage of parents who are receiving income support benefits from the government.

II. Data

The data used in this study are from the Integrated Data Infrastructure (IDI) provided by Statistics New Zealand (Stats NZ). The IDI is a large research database which contains information about people and communities in the areas of education and training, income and work, benefits and social services, demographic information, tax, health, justice, housing etc. Data is compiled with the help of different government agencies and ministries, surveys conducted by Stats NZ, and some non-government organisations as well (refer to Figure 2).

The process of getting access to IDI is very well designed and organised. Stats NZ have set up secure data labs in different cities throughout New Zealand. Researchers who require access to the data need to go through a comprehensive application and training process. Specifically, a researcher has to first apply to get access to the data, providing a research proposal with a list of variables required. Stats NZ check this research proposal in detail, along with the applicant's CV and reports from two referees. Once a proposal is approved, the researcher has to go through a confidentiality-training programme. The whole process usually takes at least two months from data application to access.

The data used in this study contain information on each student who graduated or left a NZ secondary school between 2009 and 2016. For our analysis, we use the variables shown in Table 1.

Recall that for each student in high school (where we measure NCEA and UE achievement), we need to refer back to his/her fifth birthday and hence access to primary school education. As information about actual primary school enrolment date is very sparse for our older cohort (who turned five sometime between 1990 and 2000), this study uses *potential* enrolment (please refer back to

Figure ***1***) instead of actual enrolment in school. More importantly, due to parents' choice in timing the start of school of their children (between 5 and 6 years of age), actual enrolment is likely to suffer from endogeneity. Potential enrolment would need to be used as an instrumental variable if actual enrolment were available. We use an intent-to-treat approach in its absence.

III. Methods and Results

Table 2 descriptively shows the characteristics of students in our sample. The Ministry of Education data in the IDI contains around 541,455 records on high school leavers. The first restriction we made was restricting the sample to those who left school because they had finished school (i.e. 'end of schooling') as we do not want to include students leaving school for other purposes such as to continue studies elsewhere in New Zealand or abroad. The second restriction was to isolate only domestic students[5] as we only want students who started and finished school in New Zealand. Then, we checked for duplicate observations in the data set and found 97 duplicate observations. We kept the latest record for each individual determined by comparing the students' recorded age, highest NCEA level, school leaving year, and the latest address. We also checked for inconsistencies (e.g., a student with more than one gender recorded, students with abnormal dates of birth) and removed those individuals. After all the restrictions, we were left with 411,765 observations.[6]

The population for our key analysis (Table 4-6) is somewhat different than the publically available data provided by Ministry of Education on the Education Counts website[7] for two reasons: 1. we restrict our analysis to domestic students only; and 2. we include non-NCEA classification systems such as International Baccalaureate in our model by converting them to NCEA equivalent levels (refer to Appendix Table A1). For a more detailed description of the difference in population, refer to Appendix Table A2.

We use NCEA and UE achievement as outcome measures in our analysis. The exact date of birth would be ideal for the construction of the key explanatory variable, 'the maximum

[5] We identified domestic, New Zealand-born students using two different variables: i) One on the type of student - domestic, exchange and international fee paying and ii) the other on refugee status - New Zealand born, refugee, or migrant. We focus on 'domestic' and 'New Zealand born' students in our main analysis.

[6] All the numbers of observations reported here are very close to the exact values but not exactly the same. We do not report the exact numbers of observations because of the Stats NZ privacy clause.

[7] https://www.educationcounts.govt.nz/statistics/schooling/senior-student-attainment/school-leavers2

length of schooling' but unfortunately is not available in the IDI data set provided by Stats NZ.[8] In its absence, we randomly created the date of birth for each student based on information about his/her month (and year) of birth and we calculated the maximum length of time spent in school accordingly. Later, we check the robustness of our results by re-estimating our models for alternative dates of birth: the 1st, 15th, and last of each month. The results (in Table ***2***9) show that there is no substantial impact of this on our results.

III.A. Exogeneity of the Maximum Length of Schooling

Crucial to our analysis is the assumption that children's birth dates are not timed with the school calendar in mind and that parental characteristics do not systematically differ among children born at different points during the school year. In other words, we assume that the timing of the fifth birthday, and hence the maximum length of schooling, are exogenous. To test this, we estimate the following model:

*Maximum length of time spent in school = f(age, age^2, female, ethnicity, school region, year of birth, school decile, school fixed effects, year*region)*

As reported in Table 1, the *maximum length of time spent in school* measures the amount of time spent in Years 0 and 1 of school (in months), *age* is the student's age at the start of Year 2, and *school decile* is secondary school deprivation decile used as a proxy for socioeconomic characteristics of the student's community. The standard errors are corrected for clustering at the school level and are robust to heteroscedasticity. Table 3 shows the results.

To suggest that our results are exogenous, we expected to find no significance in any of the variables apart from age (which is closely, and mechanically, linked with the maximum length of time spent in school). Consistent with our hypothesis, the exogeneity check shows the significance of age and age squared. Surprisingly, the results also suggest that being Māori decreases the *potential* amount of time spent in school compared to being New Zealand European. However, the effect is minute; Māori children spend 0.003 less months – or 0.09 of a day less – in school than New Zealand Europeans.

Based on this analysis, it is reasonable to assume that parents are not trying to time birth based on school start dates five years later.

[8] Due to the privacy clause of data in the IDI, Stats NZ do not provide the exact date of birth of students to prevent revealing the identity of any individual. There are no exceptions to this rule.

III.B. The Effects of the Time Spent in School on Later Educational Outcomes

We move next to our key analysis of the influence of the time spent in school on later educational outcomes measured by NCEA and UE results. We estimate four separate regressions:

1. NCEA 1 = At least NCEA level 1 achieved (Table 4);
2. NCEA 2 = At least NCEA level 2 achieved (Table 5);
3. NCEA 3 = NCEA level 3 achieved (Table 6);
4. UE achieved (Table 7).

The models have the following specification:

*NCEA1/NCEA2/NCEA3/UE = f(maximum length of time spent in school, age, age^2, female, ethnicity, school region, year of birth, school decile, school fixed effects, year*region)*

We use two different estimators: a linear probability model (LPM) and a probit. Because the probit (and logit) models did not converge with the inclusion of school fixed effects and the large number of year*region interactions, we focus on LPM in our main analysis. The results of a more parsimonious probit model are available in the Appendix (Table A3-A6) and are similar to the LPM results. All regressions control for age and age squared, a gender dummy, six ethnicity dummies, thirteen year of birth dummies, fifteen region dummies, year of birth and region interaction dummies, nine school deprivation decile dummies and school fixed effects. Standard errors are corrected for clustering at the school level and are robust to heteroscedasticity.

An important thing to note is that the sample size for all four regressions is the same. For instance, even if some students drop out of school after achieving NCEA level 1, they are still part of our analysis for NCEA level 2, NCEA level 3, and UE and are considered as students who have not achieved these levels (please refer to the Appendix Figure A2).

The results in Table 4 show that an additional month of the ‘maximum time spent in school’ results in an increase in achieving at least NCEA level 1 by 2 percentage points. This corresponds to about a 2.2% increase from the 89% baseline. Comparing the two extremes, being born in June rather than May increases the probability of achieving NCEA level 1 or above by 13.2 percentage points or 14.8%.[9] The ethnicity dummies show patterns similar to

[9] This is calculated by multiplying the effects of 1 month by 6.6; i.e., the difference between the maximum and the minimum potential time spent in school.

those found in the previous literature (Tofi, Flett, & Timutimu-Thorpe, 1996; Nakhid, 2003; Anae, Anderson, Benseman, & Coxon, 2002). On average, Māori and Pacific students are less likely to achieve NCEA level 1 than Asian or New Zealand European students.

For NCEA level 2 (Table 5), an additional month of the 'maximum time spent in school' results in a 3.5 percentage point increase in achievement from the 84% baseline. This makes it about a 4.2% increase. Comparing the two extremes as before, being born in June rather than May increases the probability of achieving NCEA level 2 or above by 23.1 percentage points or 27.5%.

We see that there is an increasing impact of the 'maximum length of schooling' as we move up the NCEA levels. For NCEA level 3 (Table 6), an additional month of the 'maximum time spent in school' results in an increase in achievement by 3.9 percentage points, or 6.2% compared to a sample mean baseline at 63%. Again, comparing the two extremes, being born in June rather than May increases the probability of achieving NCEA level 3 by 25.7 percentage points or 40.9%.

The effects of early schooling on NCEA level 3 and UE are very similar. This is not too surprising given the importance of NCEA level 3 credits in being awarded UE. An additional month of the 'maximum time spent in school' results in an increase of 2.2 percentage points in the achievement of UE, compared to a sample mean baseline of 43% (Table 7). This implies about a 5.2% increase; as compared to a 6.2% increase for NCEA level 3. When comparing the two extreme cases, being born in June rather than May increases the probability of achieving UE by 14.5 percentage points or 34.2%.

IV. Credibility and Heterogeneity of Results

In this section, we subject our main results to a series of robustness and falsification checks. We also investigate whether the effects are homogeneous across socio-demographic groups or whether they are concentrated in certain sub-populations. In particular, we extend the previous analysis here in the following ways:

Robustness Checks

- Excluding students born in May or June (Table 8)
- Alternative date of birth assumptions: the 1^{st}, 15^{th}, and last of each month (Table 9)

Falsification Checks

- Placebo group 1: migrants and refugees (Table 10)
- Placebo group 2: international fee paying students (Table 10)

Heterogeneity

- By gender (Table 11)
- By ethnicity (Table 12)
- By school decile group (Table 13)

IV.A. Robustness Checks

IV.A.1. Excluding Students Born in May or June

Our analysis relies on information about the students' date of birth which is used to determine the potential time spent in school prior to Year 2. The biggest difference in early formal education then theoretically occurs between children born in May and those born in June. Comparing the two extremes, a child born in May could spend 10.8 months less in school than a child born in June. However, in these extreme cases, the correspondence between *potential* schooling and *actual* schooling is likely to be the weakest. For example, schools or parents often suggest placing a child born in May into Year 0, not directly Year 1. Removing students born in May or June from the analysis (Table 8) does not qualitatively change our main findings.

IV.A.2. Alternative Proxies for the Date of Birth

As mentioned above, the exact date of birth is unfortunately unavailable in the IDI data set (only month and year of birth are) and we have assigned dates of birth randomly within each month. To test the sensitivity of our results to this 'noise', we have re-estimated all of our models using alternative assumptions about the exact date of birth, assigning to each individual the 1^{st}, 15^{th}, or last day of each month (Table 9). As the imputed potential length of schooling decreases (e.g., the 15^{th} vs. the 1^{st}), holding later outcomes constant, we would expect to estimate higher returns per month. This is indeed what Table 9 shows. However, qualitatively, the main results withstand this robustness check.

IV.B. Falsification Checks

Our main analysis restricts the sample to 'domestic' and 'New Zealand born' students. This is done in an effort to exclude students who started primary school abroad (under a different school start policy) but later participated in NCEA/UE assessment in New Zealand and hence appear in our high school dataset. We would not expect such students to benefit from being born later in the year and making use of early formal schooling in Year 0. However, while unsuitable for our main analysis, students born (and educated) overseas can help verify the credibility of our baseline model in a placebo test. In particular, our results gain credibility if they hold for domestic students but do *not* hold for migrants or international students.

IV.B.1. Placebo Group 1: Migrants and Refugees

The IDI dataset reports the migration status of each individual, distinguishing between: migrants, refugees, and New Zealand born students. In our first placebo test, we focus on migrants and refugees. As expected, the ethnic composition of this group is diverse with 15% Indian, 15% Chinese, 13% Samoan, 8% Japanese, and many other smaller groups. As a whole, the migrant and refugee community is represented about equally across high school deciles and achieves results comparable to New Zealand born students (e.g., 46% vs. 43% achieving UE, respectively).

The placebo test (middle column of Table 10) strongly suggests that our main results are not spurious. Specifically, as expected, the migrant/refugee community does not benefit from the New Zealand primary school start policy in the way that domestic students do. All of the estimated coefficients are close to zero and many have a negative sign.

IV.B.2. Placebo Group 2: International Fee Paying Students

The population for our second placebo test consists of international fee paying students. These students are not New Zealand born and they are also not classified as either migrants or refugees. A large majority of them come from Asia: 47% are Chinese, 15% Korean, and 8% Japanese. These students tend to attend high decile schools, with around 75% in deciles 7-10. Their achievement is on average comparable to New Zealand born students (e.g., 38% vs. 43% achieving UE, respectively).

To our surprise, the second placebo test (last column of Table 10) produces results similar to the baseline model for domestic, New Zealand born students. We intend to explore this puzzle further in the near future.

IV.C. Heterogeneity of Effects

Next, we explore whether the beneficial effects of early schooling occur broadly or whether they are more concentrated in certain socio-demographic groups. Specifically, we investigate the potential heterogeneity of effects by: gender, ethnicity, and school decile.

With respect to gender (Table 11), we observe large and positive effects in both groups, but especially among male students who experience larger benefits in absolute terms as well as relative to their (lower) mean performance.

Comparing by ethnicity (Table 12), early school attendance seems to have the largest benefits for Māori students, followed by New Zealand Europeans, and – only weakly - Asians.

Finally, our results by school decile group (Table 13) point to a non-monotonic relationship between socio-economic disadvantage and the benefits of early formal schooling. In particular, returns to early education seem moderate among decile 1-4 as well as decile 8-10 students. On the other hand, students in decile 5-7 schools experience large benefits, especially at the highest level of achievement as measured by NCEA level 3 and UE. One interpretation of these findings is that low decile schools provide valuable early formal education but are constrained by own resources and/or (the lack of) parental effort to complement/endorse school activities at home (Ali and Menclova, 2018). At the other end of the spectrum, children from high decile schools may be using the school environment and in-home learning as substitutes (Leuven et al., 2010).

V. Conclusion

Due to the distinctive schooling system of New Zealand, in which children can begin school right after their fifth birthday, we were able to evaluate the effect of the potential length of schooling on NCEA and UE results, autonomous from the effect of age. Controlling for demographic and socio-economic characteristics, we find that increasing the maximum length of schooling substantially increases the probability of achieving NCEA and UE results.

The magnitudes are shown in Table 14 (where the 6.6 months category illustrates the two extreme cases, i.e., being born in June rather than May). Controlling for demographic and socio-economic characteristics, we find that an additional month spent in Years 0/1 increases the probability of achieving NCEA level 1 by 2%, NCEA level 2 by 4%, NCEA level 3 by 6%, and UE by 5%. Thus, differences in the timing of birth – and hence in school start – seem to have large effects on achievement even years later, in high school.

References

Ali, A. I., & Menclova, A. (2018). Returns to Early Formal Education: A Replication of Leuven et al. (2010). Working Papers in Economics 18/19, University of Canterbury, Department of Economics and Finance.

Anae, M., Anderson, H., Benseman, J., & Coxon, E. (2002). Pacific peoples tertiary education: Issues of participation. Ministry of Education. Wellington: Auckland Uniservices Limited.

Angrist, J. D., & Krueger, A. B. (1992). The Effect of Age at School Entry on Educational Attainment: An Application ofInstrumental Variables with Moments from Two Samples. Journal of the American Statistical Association , 87 (418), 328-336.

Barnsley, R. H., & Thompson, A. H. (1988). Birthdate and success in minor hockey: The key to the NHL. Canadian Journal of Behavioural Science , 20 (2), 167-176.

Barnsley, R. H., Thompson, A. H., & Barnsley, E. P. (1985). Hockey success and birthdate: The relative age effect. Journal of the Canadian Association of Health, Physical Education and Recreation , 51.

Boucher, J. L., & Mutimer, B. T. (1994). The relative age phenomenon in sport: A replication and extension with ice-hockey players. Research Quarterly for Exercise and Sport , 65 (4), 377-381.

Cahan, S., & Cohen, N. (1989). Age versus schooling effects on intelligence development. Child Development , 60, 1239-149.

Cliffordson, C., & Gustafsson, J.-E. (2008). Effects of age and schooling on intellectual performance: Estimates obtained from analysis of continuous variation in age and length of schooling. Intelligence , 36 (2), 143-152.

Cobley, S., Baker, J., Wattie, N., & McKenna, J. (2009). Annual Age-Grouping and Athlete Development: A Meta-Analytical Review of Relative Age Effects in Sport. Sports Medicine , 39 (3), 235-256.

Crawford, C., Dearden, L., & Greaves, E. (2013). When You Are Born Matters: The Impact of Date of Birth on Educational Outcomes in England. London: Institute for Fiscal Studies (IFS).

Daniel, T. E., & Janssen, C. T. (1987). More on the relative age effect. Canadian Association for Health, Physical Education, and Recreation , 53, 21-24.

Datar, A. (2004). Does delaying kindergarten entrance give children a head start? Economics of education review , 25, 43-62.

DiPasquale, G. W., Moule,, A., & Flewelling, R. (1980). The Birthdate Effect. Journal of Learning Disabilities , 13, 234-248.

Fumarco, L., Gibbs, B. G., Jarvis, J. A., & Rossi, G. (2017). The relative age effect reversal among the National Hockey League elite. Plos One , 12 (8).

Grondin, S., Deshaies, P., & Nault, L. (1984). Trimestres de naissance et participation au hockey et au volleyball. La Revue Quebecoise de l'Activite Physique , 2, 97-103.

Kawaguchi, D. (2011). Actual age at school entry, educational outcomes, and earnings. Journal of the Japanese and International Economies , 25 (2), 64-80.

Leuven, E., Lindahl, M., Oosterbeek, H., & Webbink, D. (2010). Expanding schooling opportunities for 4-year-olds. Economics of Education Review , 29 (3), 319-328.

Lleras-Muney, A. (2005). The Relationship Between Education and Adult Mortality in the United States. Review of Economic Studies , 72, 189-221.

Lubotsky, D., & Kaestner, R. (2016). Do 'Skills Beget Skills'? Evidence on the effect of kindergarten entrance age on the evolution of cognitive and non-cognitive skill gaps in childhood. Economics of Education Review , 53, 194-206.

Musch, J., & Grondin, S. (2001). Unequal Competition as an Impediment to Personal Development: A Review of the Relative Age Effect in Sport. Developmental Review , 21, 147–167.

Nakhid, C. (2003). " Intercultural" Perceptions, Academic Achievement, and the Identifying Process of Pacific Islands Students in New Zealand Schools. Journal of Negro Education , 297-317.

New Zealand Qualifications Authority. (2013/14). New Zealand Government.

Shui, M. (2017). What grades do I need to get into a NZ university? MyTuition.

Strøm, B. (2004). Student achievement and birthday effects. Schooling and Human Capital in the Global Economy: Revisiting the Equity-Efficiency Quandary. Munich: CESifo-Harvard University/PEPG Conference.

Tofi, T., Flett, R., & Timutimu-Thorpe, H. (1996). Problems faced by pacific island students at university in New Zealand: some effects on academic performance and psychological wellbeing. New Zealand Journal of Educational Studies , 31 (1), 51-59.

Warren, C. W., Levin, M., & Tyler, C. (1986). Season of Birth and Academic Achievement. Educational and Psychological Research , 6, 111-123.

Zhang, S., Zhong, R., & Zhang, J. (2017). School starting age and academic achievement: Evidence from China's junior high schools. China Economic Review , 44, 343-354.

Figure 1. The relationship between the maximum length of time spent in school and the date of birth for a given cohort

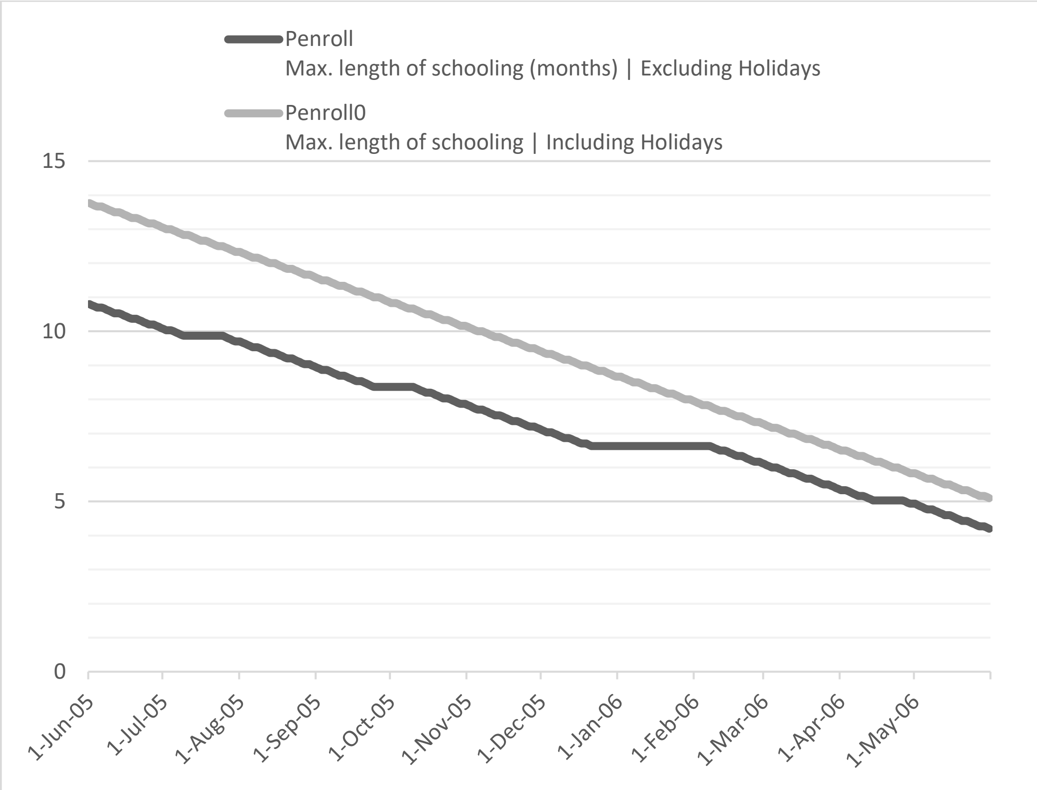
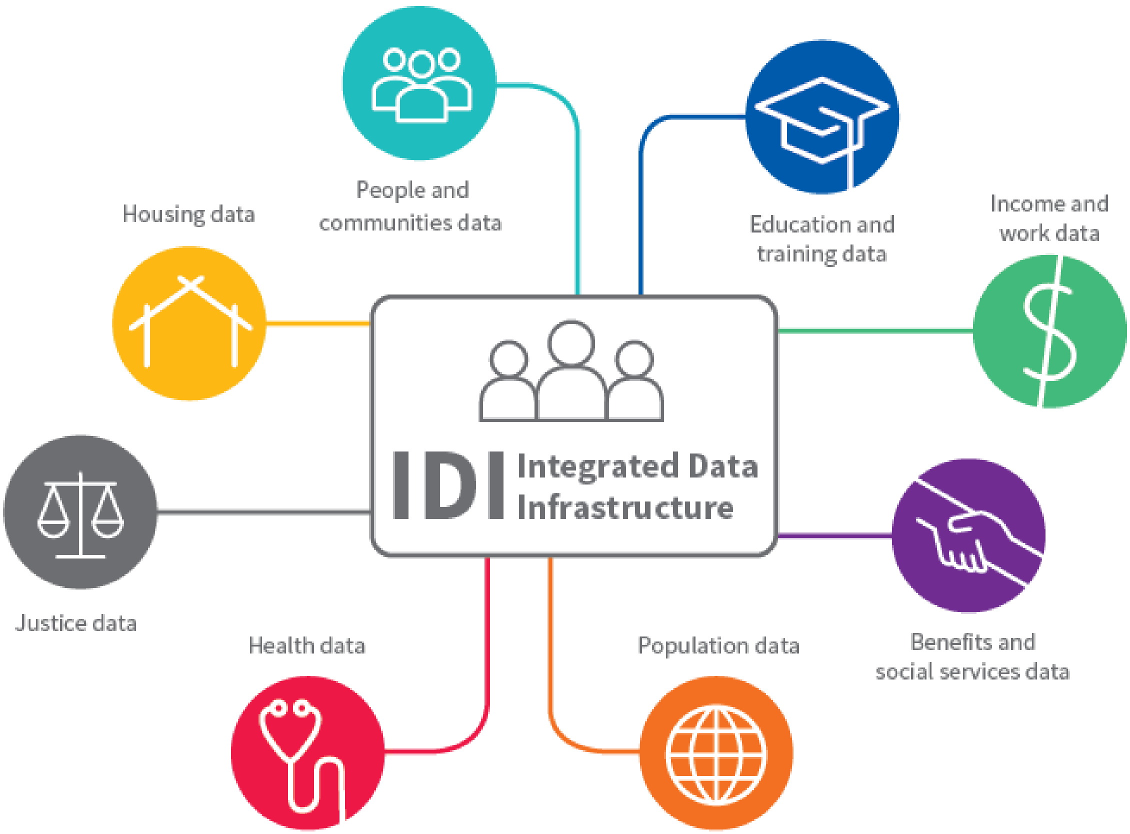


Figure 2. Data in the Integrated Data Infrastructure (IDI)



Source: (http://archive.stats.govt.nz/browse_for_stats/snapshots-of-nz/integrated-data-infrastructure.aspx), Integrated Data Infrastructure, Statistics New Zealand. Accessed on 28th November, 2018.

Table 1. Description of variables

Name of Variable	Details
NCEA 1	At least NCEA level 1 achieved *(0/1)*
NCEA 2	At least NCEA level 2 achieved *(0/1)*
NCEA 3	NCEA level 3 achieved *(0/1)*
UE	University Entrance achieved *(0/1)*
Maximum length of time spent in school	Potential enrolment in months (time spent in school) without holidays – based on a random selection of birth date within a given month
Age m	Age in months at the start of Year 2 of school
Age m^2	Age in months - squared at the start of Year 2 of school
Female	Gender of the student *(0/1)*
Ethnicity	Ethnicity of the individual *(New Zealand European, Māori, Australian, European, Pacific People, Asian, Other ethnicity, Not stated)*
Dob y	Year of birth *(1988 to 2001)*
School decile	School deprivation decile *(1-10)*
School region	The region of the school (*Northland, Auckland, Waikato, Bay of Plenty, Gisborne, Hawkes Bay, Taranaki, Manawatu-Whanaganui, Wellington, West Coast, Canterbury, Otago, Southland, Tasman, Nelson, Marlborough*)

Table 2. Descriptive Statistics (mean values and standard deviations)

		Male	Female	NZ European	Māori	Asian	Australian	European	Pacific People	Total
NCEA 1	*No. of observations*	163,458	167,862	206,310	58,737	21,867	2,028	14,916	21,471	331,320
	Mean	0.868	0.911	0.920	0.753	0.965	0.890	0.950	0.853	0.890
	Standard Deviation	0.339	0.284	0.271	0.432	0.185	0.312	0.218	0.355	0.313
NCEA 2	*No. of observations*	163,458	167,862	206,310	58,737	21,867	2,028	14,916	21,471	331,320
	Mean	0.809	0.870	0.872	0.673	0.953	0.845	0.919	0.806	0.840
	Standard Deviation	0.393	0.336	0.334	0.469	0.212	0.362	0.273	0.396	0.367
NCEA 3	*No. of observations*	163,458	167,862	206,310	58,737	21,867	2,028	14,916	21,471	331,320
	Mean	0.554	0.702	0.661	0.402	0.879	0.658	0.760	0.570	0.629
	Standard Deviation	0.497	0.457	0.473	0.490	0.327	0.475	0.427	0.495	0.483
UE	*No. of observations*	210,246	201,522	248,667	80,163	24,606	2,511	17,430	31,023	411,765
	Mean	0.356	0.496	0.478	0.185	0.743	0.473	0.589	0.250	0.425
	Standard Deviation	0.479	0.500	0.500	0.388	0.437	0.499	0.492	0.433	0.494
Maximum length of time spent in school	*No. of observations*	210,246	201,522	248,670	80,163	24,606	2,511	17,430	31,023	411,765
	Mean	7.461	7.465	7.468	7.461	7.438	7.432	7.449	7.448	7.463
	Standard Deviation	1.807	1.811	1.804	1.820	1.785	1.841	1.828	1.824	1.809
Age m	*No. of observations*	210,246	201,522	248,670	80,163	24,606	2,511	17,430	31,023	411,765
	Mean	74.125	74.133	74.140	74.124	74.086	74.064	74.093	74.094	74.129
	Standard Deviation	S	S	S	S	S	S	S	S	S
Female	*No. of observations*	-	-	248,670	80,163	24,606	2,511	17,430	31,023	411,765
	Mean	-	-	0.488	0.491	0.495	0.484	0.483	0.498	0.489
	Standard Deviation	-	-	0.500	0.500	0.500	0.500	0.500	0.500	0.500
Dob y	*No. of observations*	210,246	201,522	248,670	80,163	24,606	2,511	17,430	31,023	411,765
	Mean	1995	1995	1995	1995	1995	1995	1995	1995	1995
	Standard Deviation	S	S	S	S	S	S	S	S	S
School decile	*No. of observations*	200,721	189,072	234,972	74,925	23,805	2,379	16,578	30,426	389,796
	Mean	6.127	6.137	6.783	4.405	6.873	6.991	7.638	3.803	6.132
	Standard Deviation	2.611	2.660	2.249	2.460	2.584	2.381	2.081	3.483	2.635

Note: All figures have been randomly rounded to base 3 (RR3) – the number is randomly rounded to either the nearest base above or below the number – following the Stats NZ privacy requirement. Standard Deviations for age m, age y and dob y have been supressed (S) due to a privacy clause. ‘Age m’ is Age in months at the start of Year 2

.

Table 3. Exogeneity Check

Linear regression		Number of observations	391,368
		R-squared	0.9918
		Root MSE	0.16364
(Std. Err. adjusted for 543 clusters by school)			

Maximum length of time spent in school	**Coef.**	**Robust Std. Err.**
Age m	-0.481***	0.004
Age m^2	0.673***	0.002
Female	0.001	0.001
Ethnicity: Māori	-0.003***	0.001
Ethnicity: Asian	0.000	0.001
Ethnicity: Australian	-0.003	0.003
Ethnicity: European	0.003	0.001
Ethnicity: Pacific People	0.001	0.001

Note: All regressions include thirteen year of birth dummies, eighteen region dummies and their interactions; nine school decile dummies; and a full set of school fixed effects. *, **, and *** indicate statistical significance at 95%, 99%, and 99.9% confidence levels respectively.

Table 4. Effect of maximum schooling on achieving at least NCEA level 1

Linear regression		Number of observations	315,969
		R-squared	0.1669
		Root MSE	0.26645
(Std. Err. adjusted for 542 clusters by school)			

NCEA 1	**Coef.**	**Robust Std. Err.**
Maximum length of time spent in school	0.020***	0.003
Age m	0.035***	0.006
Age m^2	-0.031***	0.005
Female	0.048***	0.002
Ethnicity: Māori	-0.121***	0.006
Ethnicity: Asian	0.016***	0.003
Ethnicity: Australian	-0.029***	0.007
Ethnicity: European	0.009***	0.003
Ethnicity: Pacific People	-0.049***	0.007

Note: All regressions also include thirteen year of birth dummies, eighteen region dummies, year of birth and region interactions, nine school decile dummies, and school fixed effects. *, **, and *** indicate statistical significance at 95%, 99%, and 99.9% confidence levels respectively.

Table 5. Effect of maximum schooling on achieving at least NCEA level 2

Linear regression	Number of observations	315,969
	R-squared	0.1738
	Root MSE	0.31662
(Std. Err. adjusted for 542 clusters by school)		

NCEA 2	Coef.	Robust Std. Err.
Maximum length of time spent in school	0.035***	0.004
Age m	0.039***	0.007
Age m^2	-0.040***	0.006
Female	0.069***	0.003
Ethnicity: Māori	-0.145***	0.006
Ethnicity: Asian	0.037***	0.004
Ethnicity: Australian	-0.030***	0.008
Ethnicity: European	0.020***	0.003
Ethnicity: Pacific People	-0.051***	0.007

Note: All regressions also include thirteen year of birth dummies, eighteen region dummies, year of birth and region interactions, nine school decile dummies, and school fixed effects. *, **, and *** indicate statistical significance at 95%, 99%, and 99.9% confidence levels respectively.

Table 6. Effect of maximum schooling on achieving NCEA level 3

Linear regression	Number of observations	315,969
	R-squared	0.2178
	Root MSE	0.4232
(Std. Err. adjusted for 542 clusters by school)		

NCEA 3	Coef.	Robust Std. Err.
Maximum length of time spent in school	0.039***	0.005
Age m	0.023*	0.010
Age m^2	-0.031***	0.008
Female	0.161***	0.005
Ethnicity: Māori	-0.181***	0.005
Ethnicity: Asian	0.137***	0.008
Ethnicity: Australian	-0.010	0.011
Ethnicity: European	0.058***	0.005
Ethnicity: Pacific People	-0.094***	0.009

Note: All regressions also include thirteen year of birth dummies, eighteen region dummies, year of birth and region interactions, nine school decile dummies, and school fixed effects. *, **, and *** indicate statistical significance at 95%, 99%, and 99.9% confidence levels respectively.

Table 7. Effect of maximum schooling on UE

Linear regression	Number of observations	391,368
	R-squared	0.2464
	Root MSE	0.43137
(Std. Err. adjusted for 543 clusters by school)		

UE	Coef.	Robust Std. Err.
Maximum length of time spent in school	0.022***	0.005
Age m	-0.002	0.009
Age m^2	-0.007	0.007
Female	0.147***	0.004
Ethnicity: Māori	-0.202***	0.005
Ethnicity: Asian	0.179***	0.009
Ethnicity: Australian	-0.014	0.010
Ethnicity: European	0.064***	0.006
Ethnicity: Pacific People	-0.194***	0.010

Note: All regressions also include thirteen year of birth dummies, eighteen region dummies, year of birth and region interactions, nine school decile dummies, and school fixed effects. *, **, and *** indicate statistical significance at 95%, 99%, and 99.9% confidence levels respectively.

Table 8. Effect of maximum schooling: without students born in May or June

		Main Model	w/o May & June
NCEA 1	*Coefficient*	0.020***	0.011**
	Standard error	(0.003)	(0.003)
	Percentage change	2.2%	1.2%
NCEA 2	*Coefficient*	0.035***	0.026***
	Standard error	(0.004)	(0.004)
	Percentage change	4.2%	3.1%
NCEA 3	*Coefficient*	0.039***	0.034***
	Standard error	(0.005)	(0.005)
	Percentage change	6.2%	5.4%
UE	*Coefficient*	0.022***	0.023***
	Standard error	(0.005)	(0.005)
	Percentage change	5.2%	5.4%
Number of observations	NCEA	315,969	264,138
	UE	391,368	327,201

Note: All regressions also include thirteen year of birth dummies, eighteen region dummies, year of birth and region interactions, nine school decile dummies, and school fixed effects. *, **, and *** indicate statistical significance at 95%, 99%, and 99.9% confidence levels respectively.

Table 9. Effect of maximum schooling: different date of birth assumptions

		Main Model	1st of each month	15th of each month	Last of each month
NCEA 1	*Coefficient*	0.020***	0.015***	0.023***	0.029***
	Standard error	(0.003)	(0.003)	(0.003)	(0.003)
	Percentage change	2.2%	1.7%	2.6%	3.3%
NCEA 2	*Coefficient*	0.035***	0.027***	0.038***	0.043***
	Standard error	(0.004)	(0.004)	(0.004)	(0.004)
	Percentage change	4.2%	3.2%	4.5%	5.1%
NCEA 3	*Coefficient*	0.039***	0.026***	0.041***	0.045***
	Standard error	(0.005)	(0.005)	(0.005)	(0.005)
	Percentage change	6.2%	4.1%	6.5%	7.2%
UE	*Coefficient*	0.022***	0.010	0.024***	0.030***
	Standard error	(0.005)	(0.005)	(0.004)	(0.004)
	Percentage change	5.2%	2.3%	5.7%	7.1%
Number of observations	NCEA	315,969	315,969	315,969	315,969
	UE	391,368	391,365	391,365	391,365

Note: All regressions also include thirteen year of birth dummies, eighteen region dummies, year of birth and region interactions, nine school decile dummies, and school fixed effects. *, **, and *** indicate statistical significance at 95%, 99%, and 99.9% confidence levels respectively.

Table 10. Effect of maximum schooling: placebo tests

		Main Model	Placebo group 1: Migrants and Refugees	Placebo group 2: International Fee Paying Students
NCEA 1	*Coefficient*	0.020***	0.003	0.023
	Standard error	(0.003)	(0.010)	(0.016)
	Percentage change	2.2%	0.3%	3.8%
NCEA 2	*Coefficient*	0.035***	-0.001	0.025
	Standard error	(0.004)	(0.011)	(0.017)
	Percentage change	4.2%	-0.1%	4.3%
NCEA 3	*Coefficient*	0.039***	-0.003	0.043*
	Standard error	(0.005)	(0.015)	(0.021)
	Percentage change	6.2%	-0.4%	8.7%
UE	*Coefficient*	0.022***	-0.001	0.053**
	Standard error	(0.005)	(0.015)	(0.020)
	Percentage change	5.2%	-0.2%	11.2%
Number of observations	NCEA	315,969	26,922	12,000
	UE	391,368	34,581	14,157

Note: All regressions also include thirteen year of birth dummies, eighteen region dummies, year of birth and region interactions, nine school decile dummies, and school fixed effects. *, **, and *** indicate statistical significance at 95%, 99%, and 99.9% confidence levels respectively.

Table 11. Effect of maximum schooling: by gender

		Main Model	Female students	Male students
NCEA 1	*Coefficient*	0.020***	0.016***	0.025***
	Standard error	(0.003)	(0.004)	(0.005)
	Percentage change	2.2%	1.8%	2.9%
NCEA 2	*Coefficient*	0.035***	0.023***	0.048***
	Standard error	(0.004)	(0.004)	(0.006)
	Percentage change	4.2%	2.6%	5.9%
NCEA 3	*Coefficient*	0.039***	0.035***	0.043***
	Standard error	(0.005)	(0.006)	(0.007)
	Percentage change	6.2%	5.0%	7.8%
UE	*Coefficient*	0.022***	0.020**	0.024***
	Standard error	(0.005)	(0.006)	(0.006)
	Percentage change	5.2%	4.0%	6.7%
Number of observations	NCEA	315,969	159,069	156,900
	UE	391,368	189,813	201,555

Note: All regressions also include thirteen year of birth dummies, eighteen region dummies, year of birth and region interactions, nine school decile dummies, and school fixed effects. *, **, and *** indicate statistical significance at 95%, 99%, and 99.9% confidence levels respectively.

Table 12. Effect of maximum schooling: by ethnicity

		Main Model	NZ European	Māori	Asian
NCEA 1	*Coefficient*	0.020***	0.013***	0.057***	-0.003
	Standard error	(0.003)	(0.003)	(0.011)	(0.006)
	Percentage change	2.2%	1.4%	7.6%	-0.3%
NCEA 2	*Coefficient*	0.035***	0.034***	0.065***	0.008
	Standard error	(0.004)	(0.004)	(0.011)	(0.007)
	Percentage change	4.2%	3.9%	9.7%	0.8%
NCEA 3	*Coefficient*	0.039***	0.056***	0.028*	0.012
	Standard error	(0.005)	(0.006)	(0.011)	(0.013)
	Percentage change	6.2%	8.5%	7.0%	1.4%
UE	*Coefficient*	0.022***	0.034***	0.025**	0.016
	Standard error	(0.005)	(0.006)	(0.009)	(0.015)
	Percentage change	5.2%	7.1%	13.5%	2.2%
Number of observations	NCEA	315,969	196,362	54,921	21,633
	UE	391,368	235,500	75,093	24,279

Note: All regressions also include thirteen year of birth dummies, eighteen region dummies, year of birth and region interactions, nine school decile dummies, and school fixed effects. *, **, and *** indicate statistical significance at 95%, 99%, and 99.9% confidence levels respectively.

Table 13. Effect of maximum schooling: by school decile group

		Main Model	Deciles 1-4	Deciles 5-7	Deciles 8-10
NCEA 1	*Coefficient*	0.020***	0.028**	0.025***	0.011**
	Standard error	(0.003)	(0.008)	(0.005)	(0.003)
	Percentage change	2.2%	3.4%	2.8%	1.1%
NCEA 2	*Coefficient*	0.035***	0.034***	0.046***	0.026***
	Standard error	(0.004)	(0.008)	(0.007)	(0.005)
	Percentage change	4.2%	4.5%	5.4%	2.8%
NCEA 3	*Coefficient*	0.039***	0.016	0.059***	0.036***
	Standard error	(0.005)	(0.008)	(0.009)	(0.007)
	Percentage change	6.2%	3.2%	9.8%	0.0%
UE	*Coefficient*	0.022***	0.011	0.039***	0.015
	Standard error	(0.005)	(0.007)	(0.008)	(0.008)
	Percentage change	5.2%	4.3%	9.9%	2.3%
Number of observations	NCEA	315,969	85,260	110,070	117,900
	UE	391,368	114,972	138,180	135,402

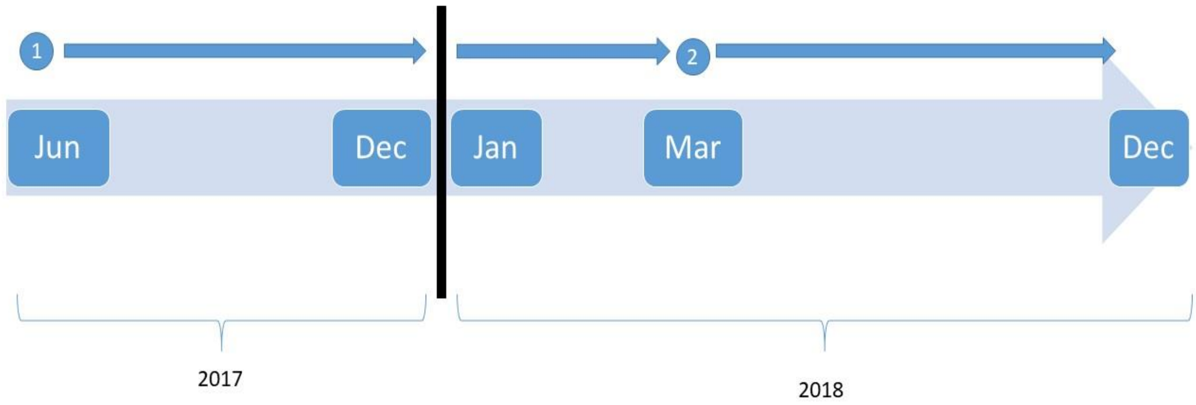
Note: All regressions also include thirteen year of birth dummies, eighteen region dummies, year of birth and region interactions, nine school decile dummies, and school fixed effects. *, **, and *** indicate statistical significance at 95%, 99%, and 99.9% confidence levels respectively.

Table 14. Effects of an increase in maximum schooling on NCEA and UE in terms of percentage points and percentages

	Sample mean	Increase in maximum schooling by:	Effects in terms of percentage points (pp) and percentages (%)
NCEA 1 (at least NCEA level 1)	89%	1 month	2.0 pp
		1 month	2.2%
		6.6 months	14.8%
NCEA 2 (at least NCEA level 2)	84%	1 month	3.5 pp
		1 month	4.2%
		6.6 months	27.5%
NCEA 3	63%	1 month	3.9 pp
		1 month	6.2%
		6.6 months	40.9%
UE	43%	1 month	2.2 pp
		1 month	5.2%
		6.6 months	34.2%

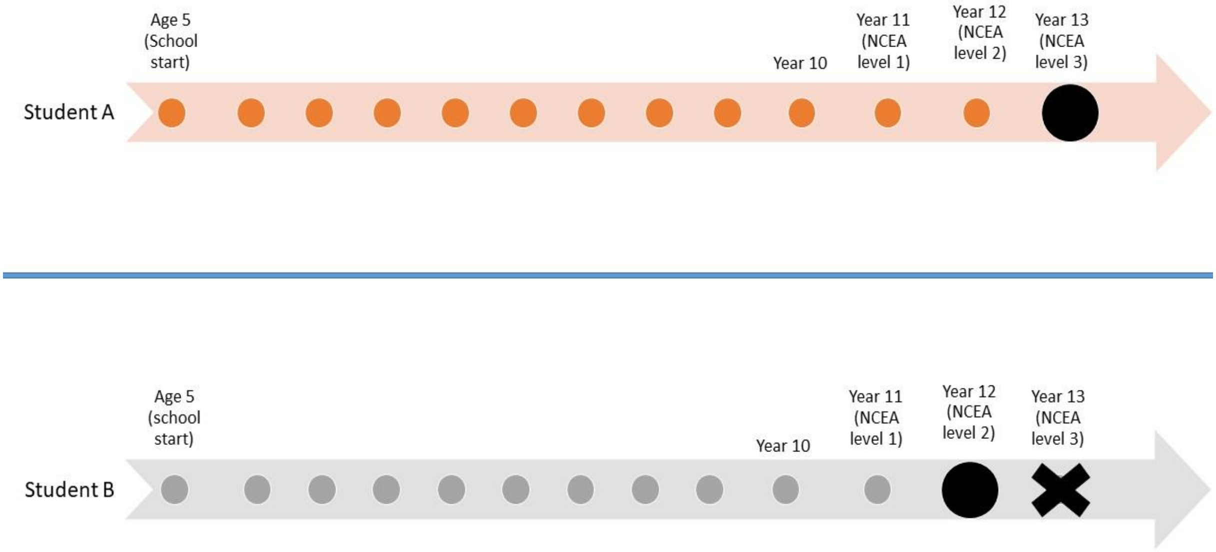
Appendix

Figure A1. Comparison of two hypothetical students starting school at different times



The diagram shows the example of two students labelled with 1 and 2 in calendar years 2017 & 2018 and how much time they will spend in school before the start of Year 2. Take student 1, who starts school in June 2017. This student will spend the rest of year 2017 (June – December) and the entire next year, 2018 (January - December) in school before he/she starts Year 2. Now take student 2, who starts school in March 2018. This student will only spend March till December of 2018 in school and will move on to Year 2 next year. Comparing the two students, student 1 has spent roughly 18 months in school whilst student 2 has spent only 10 months in school before starting Year 2. Taking this even further, another student born in May of 2018 will only spend 6 months in school before Year 2.

Figure A2. Comparison of two hypothetical students achieving different NCEA levels



The diagram above gives an example of two different students (Student A and Student B) each achieving different NCEA levels. The small dots show the time spent in school. The large black dot shows the last year of school or last NCEA achievement level. The black cross shows the year or NCEA level not attended/not achieved. If we compare the two students, we can see that student A attended all 13 years and achieved NCEA level 3 whereas student B achieved only NCEA level 2 (either because he/she did not attend Year 13 or did but did not achieve NCEA level 3). When analysed in a model of NCEA level 3 achievement, both of these students are in the sample, with student A achieving level 3 and student B not achieving/attending level 3. The important thing to note here is that the sample size for different models - NCEA level 1, level 2, and level 3 (and University Entrance) is the same. In other words, drop-outs are treated as non-achievers along with those who attempted the assessment but did not succeed in it.

Table A1. Highest attainment variable – classification table

Code	Description	Percent	NCEA-Equivalent Measure
0	No Formal Attainment	2.18	NCEA not achieved
10	1 – 13 credits at level 1	1.11	NCEA not achieved
13	Other level 1 NQF Qualification	0.16	At least NCEA Level 1 achieved
14	NCEA level 1 not further defined	0.00	At least NCEA Level 1 achieved
15	NCEA level 1 achieved	3.47	At least NCEA Level 1 achieved
16	NCEA level 1 with merit	0.18	At least NCEA Level 1 achieved
17	NCEA level 1 with excellence	0.01	At least NCEA Level 1 achieved
20	1 – 13 credits at level 2	0.36	NCEA not achieved
24	NCEA level 2 not further defined	0.00	At least NCEA Level 2 achieved
25	NCEA level 2 achieved	15.77	At least NCEA Level 2 achieved
26	NCEA level 2 with merit	0.62	At least NCEA Level 2 achieved
27	NCEA level 2 with excellence	0.11	At least NCEA Level 2 achieved
30	1 – 13 credits at level 3	0.13	At least NCEA Level 1 achieved
33	Other level 3 NQF Qualification	0.49	NCEA Level 3 achieved
34	NCEA level 3 not further defined	0.00	NCEA Level 3 achieved
35	NCEA level 3 achieved	30.27	NCEA Level 3 achieved
36	NCEA level 3 with merit	12.23	NCEA Level 3 achieved
37	NCEA level 3 with excellence	4.65	NCEA Level 3 achieved
4	Other level 2 NQF Qualification	0.30	At least NCEA Level 2 achieved
40	3+ NZ Scholarships subjects	0.49	NCEA Level 3 achieved
43	National certificate at level 4	0.09	NCEA Level 3 achieved
51	14 – 39 credits at any level without level 1 literacy and numeracy credits	2.68	NCEA not achieved
52	14 – 39 credits at any level including level 1 literacy and numeracy credits	0.31	NCEA not achieved
53	40+ credits at any level without level 1 literacy and numeracy credits	2.21	NCEA not achieved
54	40+ credits at any level including level 1 literacy and numeracy credits	1.83	-----
55	30+ credits at level 2 or above	6.34	-----
56	30+ credits at level 3 or above	11.37	-----
60	International Baccalaureate Year 11	0.00	At least NCEA Level 1 achieved
61	International Baccalaureate Year 12	0.01	At least NCEA Level 2 achieved
62	International Baccalaureate Year 13	0.49	NCEA Level 3 achieved
70	Cambridge International Exams Year 11	0.04	At least NCEA Level 1 achieved
71	Cambridge International Exams Year 12	0.16	At least NCEA Level 2 achieved
72	Cambridge International Exams Year 13	1.87	NCEA Level 3 achieved
80	Accelerated Christian Education Year 11	0.02	At least NCEA Level 1 achieved
81	Accelerated Christian Education Year 12	0.01	At least NCEA Level 2 achieved
82	Accelerated Christian Education Year 13	0.02	NCEA Level 3 achieved
90	Other Overseas Awards Year 11	0.00	At least NCEA Level 1 achieved
91	Other Overseas Awards Year 12	0.00	At least NCEA Level 2 achieved
92	Other Overseas Awards Year 13	0.00	NCEA Level 3 achieved

Table A2. Difference in population of MOE and our analysis

	A	B	C
	MOE	Our analysis: Missing NCEA excluded	Our analysis: Missing NCEA included as NCEA not achieved
NCEA not achieved (0)	14%	11%	28%
At least NCEA 1 achieved (1)	86%	89%	72%

The table matches our data (columns B & C) with the MOE's publicly available data (column A) on the Education Counts website[10] and shows two different possibilities of using the missing values for the NCEA variable in our data. There are around 80,500 missing values for NCEA in our data set, which is approximately 19% of the total population. The table shows the difference between two scenarios: i) excluding these missing values from our analysis (column B) and ii) adding them to the category of NCEA not achieved (column C). It is evident that considering these values as missing and dropping them (column B) is more similar to the MOE's publicly available data (column A) so we have chosen this approach.

[10] https://www.educationcounts.govt.nz/statistics/schooling/senior-student-attainment/school-leavers2

Table A3. NCEA 1: OLS vs Probit

NCEA1						
Linear regression	Number of observations	315,969		**Marginal effects after probit**		
	R-squared	0.1025		y = Pr(ncea1) (predict)		
	Root MSE	0.27624		= 0.93465254		
(Std. Err. adjusted for 542 clusters by school)						
NCEA 1	**Coef.**	**Robust Std. Err.**		**NCEA 1**	**dy/dx**	**Std. Err.**
Maximum length of time spent in school	0.018***	0.003		**Maximum length of time spent in school**	0.015***	0.003
Age m	0.035***	0.007		**Age m**	0.028***	0.006
Age m^2	-0.031***	0.005		**Age m^2**	-0.025***	0.004
Female	0.042***	0.003		**Female**	0.038***	0.003
Ethnicity: Māori	-0.121***	0.007		**Ethnicity: Māori**	-0.102***	0.007
Ethnicity: Asian	0.024***	0.005		**Ethnicity: Asian**	0.029***	0.004
Ethnicity: Australian	-0.032***	0.007		**Ethnicity: Australian**	-0.039***	0.008
Ethnicity: European	0.006*	0.003		**Ethnicity: European**	0.010**	0.004
Ethnicity: Pacific People	-0.023*	0.010		**Ethnicity: Pacific People**	-0.030***	0.008

Table A4. NCEA 2: OLS vs Probit

NCEA2						
Linear regression	Number of observations	315,969		**Marginal effects after probit**		
	R-squared	0.1192		y = Pr(ncea2) (predict)		
	Root MSE	0.32654		= 0.89127716		
(Std. Err. adjusted for 542 clusters by school)						
NCEA 2	**Coef.**	**Robust Std. Err.**		**NCEA 2**	**dy/dx**	**Std. Err.**
Maximum length of time spent in school	0.035***	0.004		**Maximum length of time spent in school**	0.031***	0.004
Age m	0.040***	0.008		**Age m**	0.035***	0.008
Age m^2	-0.040***	0.006		**Age m^2**	-0.036***	0.006
Female	0.063***	0.004		**Female**	0.061***	0.004
Ethnicity: Māori	-0.142***	0.007		**Ethnicity: Māori**	-0.126***	0.007
Ethnicity: Asian	0.051***	0.006		**Ethnicity: Asian**	0.062***	0.005
Ethnicity: Australian	-0.034***	0.009		**Ethnicity: Australian**	-0.040***	0.010
Ethnicity: European	0.015***	0.004		**Ethnicity: European**	0.020***	0.005
Ethnicity: Pacific People	-0.016	0.010		**Ethnicity: Pacific People**	-0.024**	0.009

Table A5. NCEA 3: OLS vs Probit

NCEA3					
Linear regression	Number of observations R-squared Root MSE	315,969 0.1639 0.43705	**Marginal effects after probit**		
(Std. Err. adjusted for 542 clusters by school)			y = Pr(ncea3) (predict) = 0.66669332		
NCEA 3	**Coef.**	**Robust Std. Err.**	**NCEA 3**	**dy/dx**	**Std. Err.**
Maximum length of time spent in school	0.041***	0.005	**Maximum length of time spent in school**	0.046***	0.005
Age m	0.023*	0.011	**Age m**	0.028*	0.013
Age m^2	-0.032***	0.008	**Age m^2**	-0.037***	0.009
Female	0.151***	0.007	**Female**	0.169***	0.007
Ethnicity: Māori	-0.178***	0.007	**Ethnicity: Māori**	-0.186***	0.007
Ethnicity: Asian	0.162***	0.011	**Ethnicity: Asian**	0.201***	0.009
Ethnicity: Australian	-0.016	0.011	**Ethnicity: Australian**	-0.020	0.013
Ethnicity: European	0.042***	0.008	**Ethnicity: European**	0.048***	0.009
Ethnicity: Pacific People	-0.036**	0.013	**Ethnicity: Pacific People**	-0.039**	0.014

Table A6. UE: OLS vs Probit

UE					
Linear regression	Number of observations R-squared Root MSE	391,368 0.198 0.44462	**Marginal effects after probit**		
(Std. Err. adjusted for 543 clusters by school)			y = Pr(ue) (predict) = 0.42455546		
UE	**Coef.**	**Robust Std. Err.**	**UE**	**dy/dx**	**Std. Err.**
Maximum length of time spent in school	0.025***	0.005	**Maximum length of time spent in school**	0.030***	0.005
Age m	-0.003	0.010	**Age m**	-0.002	0.012
Age m^2	-0.007	0.007	**Age m^2**	-0.010	0.008
Female	0.148***	0.007	**Female**	0.173***	0.008
Ethnicity: Māori	-0.203***	0.006	**Ethnicity: Māori**	-0.230***	0.006
Ethnicity: Asian	0.209***	0.012	**Ethnicity: Asian**	0.246***	0.012
Ethnicity: Australian	-0.019	0.011	**Ethnicity: Australian**	-0.021	0.012
Ethnicity: European	0.043***	0.011	**Ethnicity: European**	0.044***	0.012
Ethnicity: Pacific People	-0.146***	0.012	**Ethnicity: Pacific People**	-0.150***	0.013

Note: All regressions in Tables A2-A6 also include thirteen year of birth dummies, eighteen region dummies, and nine school decile dummies. *, **, and *** indicate statistical significance at 95%, 99%, and 99.9% confidence levels respectively.